Biodivers Conserv
DOI 10.1007/s10531-014-0749-8

ORIGINAL PAPER

Accounting for observation reliability when inferring extinction based on sighting records

Ivan Jarić · David L. Roberts

Received: 6 April 2014 / Revised: 1 June 2014 / Accepted: 25 June 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Given the current increase in global extinction risk of species, methods that are able to estimate the probability and the time of extinction based on a sighting record of a rare or poorly studied species are becoming increasingly important. One of the major obstacles when using such methods is that many sighting records are burdened with uncertain or controversial observations. What is accepted as a valid sighting can have a substantial effect on resulting predictions. Here we present a simple modification to an existing method that allows for the inclusion of specific sighting reliabilities of individual observations into predictions of the likelihood and the time of extinction. The approach is applied to the sighting records of four presumably extinct bird species. Results indicated that the Ivory-billed Woodpecker (*Campephilus principalis*) and the Eskimo Curlew (*Numenius borealis*) may be considered as extinct, even when recent controversial sightings are included in the analysis. The Nukupu'u (*Hemignathus lucidus*) and O'ahu 'Alauahio (*Paroreomyza maculata*) could potentially still be extant, although with a low probability of persistence. The major advantage of the presented method is that, instead of applying some arbitrary threshold for the sighting reliability that is considered acceptable, it recognizes and incorporates the reliability of each observation into the resulting predictions, by estimating the most likely number of observations and the most likely time of the last observation. Its simplicity facilitates its application for assessments of a larger number species or populations, and makes it accessible as a decision tool.

Communicated by Dirk Sven Schmeller.

I. Jarić (✉)
Institute for Multidisciplinary Research, University of Belgrade, Kneza Viseslava 1, 11000 Belgrade, Serbia
e-mail: ijaric@imsi.rs; ijaric@imsi.bg.ac.rs

D. L. Roberts
Durrell Institute of Conservation & Ecology, School of Anthropology & Conservation, University of Kent, Marlowe Building, Canterbury, Kent, CT2 7NR, UK

Published online: 11 July 2014

Keywords Extinction probability · IUCN Red List · Eskimo Curlew · Ivory-billed Woodpecker · Nukupu‘u · O‘ahu ‘Alauahio

Introduction

Classifying a species as extinct or extant is essential for planned and ongoing conservation management activities, as well as for conservation practitioners when they are faced with difficult decisions in the presence of limited time and resources (Elphick et al. 2010). However, determining the fate of a rare or cryptic species can be problematic (Solow 2005). Given that there are many cases of taxa that have been considered extinct only to be later rediscovered (‘Lazarus’ taxa; Roberts and Kitchener 2006; Fisher and Blomberg 2011), conservation practitioners are usually reluctant to categorize species as extinct if there is any possibility that they could still be extant, preferring to list them as Critically Endangered (Butchart et al. 2006). Such false alarms might lead to a so-called “Romeo Error” (Collar 1998), reduce public trust in conservation management and undermine wider public support (Roberts and Kitchener 2006).

One way of addressing this problem has been through the recent development of quantitative methods that are able to test the hypothesis of a species extinction based on a sighting record (Solow 1993a, b, 2005; Burgman et al. 1995; Solow and Roberts 2003; Roberts and Solow 2003; McInerny et al. 2006; Jarić and Ebenhard 2010; Solow et al. 2012; Thompson et al. 2013; Lee et al. 2014). For rare or poorly studied species, sighting records often represent the only available data for the quantitative assessment. In such cases, these methods represent the only available quantitative approach that may provide validity and reliability to extinction assessments (Jarić and Ebenhard 2010). Moreover, in accordance with the International Union for Conservation of Nature (IUCN) criteria that call for quantitative assessments of the extinction risk when assigning threat categories (IUCN 2001), this group of methods has been advocated as a tool to estimate threat status in extant species (McCarthy 1998; Regan et al. 2000; Robbirt et al. 2006). Although the general idea of using sighting records to estimate extinction probability and the time of extinction was first introduced by paleobiologists (Strauss and Sadler 1989; Marshall 1990), it was applied for the first time within the field of conservation biology and current species extinctions by Solow (1993a, b).

One of the major obstacles when using these methods is that many sighting records are burdened with uncertain or controversial observations. It is a difficult task to decide whether certain observations of a rare or presumably extinct species are of acceptable reliability (McKelvey et al. 2008; Elphick et al. 2010; Roberts et al. 2010). Such decisions are often made *ad hoc*, with a substantial level of subjectivity involved in the process. As stated by McKelvey et al. (2008), “…the degree to which we accept or reject anecdotal data is largely a matter of belief, not reason”. However, what is accepted as a valid sighting can have a substantial impact on assessments of species extinctions (Roberts et al. 2010; Solow et al. 2012). Controversial sightings can increase the likelihood of making either a type I or type II statistical errors. Under a false assumption that the species is extinct, inappropriate management actions, or the lack of them, may lead to the actual species extinction (Jarić and Ebenhard 2010). The alternative situation may lead to costly and unnecessary sampling and management activities (Grogan and Boreman 1998). A notable example was the controversial rediscovery of the Ivory-Billed Woodpecker (Fitzpatrick

et al. 2005; Hill et al. 2006; Sibley et al. 2006), which resulted in extensive search efforts, and costly management strategies and recovery plans (McKelvey et al. 2008; Gotelli et al. 2012).

There is a need for methods that directly account for the reliability of observations, and not simply to make decisions whether to accept or reject controversial sightings. Previous studies applied Bayesian methods that take into account the reliability of controversial sightings (Solow et al. 2012; Lee et al. 2014). However, such methods are often computationally demanding. Moreover, the method proposed by Solow et al. (2012) divides the sighting period into segments with certain and controversial observations, which consequently results in all controversial observations distributed within the former period (i.e. those located within the series of certain sightings) being considered certain as well. Controversial records that fill the gaps between certain observations can also influence the predicted time of extinction, by increasing the sighting rate and therefore drawing the inferred date of extinction closer to the last sighting (Roberts et al. 2010).

In this paper, we present a simple modification of an existing method, that of Solow (1993a), which allows for inclusion of specific sighting reliabilities of individual observations into predictions of the likelihood and the time of extinction. The approach is applied to sighting records of the four presumably extinct bird species (Roberts et al. 2010): the Eskimo Curlew (*Numenius borealis*), Ivory-billed Woodpecker (*Campephilus principalis*), Nukupu'u (*Hemignathus lucidus*) and O'ahu 'Alauahio (*Paroreomyza maculata*).

Methods

Model description

To present the general approach of introducing observation reliability considerations within the existing model that estimates extinction probabilities based on sighting records, we applied it to the first and the most widely employed method, the Solow equation (Solow 1993a). Solow equation tests the hypothesis of a species presence in relation to the number of time units in which the species was recorded (n) within the whole sighting period (T):

$$p(T_n \leq t_n | T_E = T) = \left(\frac{t_n}{T}\right)^n \tag{1}$$

where T_E represents the time of extinction and t_n is the time of the last positive sighting within the observation period, with the sighting records ordered from the earliest to the latest, $t_1 < t_2 < \ldots < t_n$. The Solow (1993a) method also comprises a uniform minimum variance unbiased estimate of the extinction time, as well as the estimation of the upper bound (T_{CI}) of a $1 - \alpha$ confidence interval ($\alpha = 0.05$) for T_E (Solow 2005):

$$T_E = \frac{n+1}{n} t_n \tag{2}$$

$$T_{CI} = \frac{t_n}{\alpha^{(1/n)}} \tag{3}$$

Standard use of this group of equations treats observation records as a binary sequence of presences and absences (Gotelli et al. 2012). The approach proposed here modifies their use so that the species presences are expressed as probabilities that represent reliability of

each observation. If the reliability of the sighting record is not certain, a probability value can be assigned to each of the sightings (i.e., p_1, p_2,… p_n), indicating the likelihood that the given observation is true. Probability value close to 1 would therefore correspond to a certain observation (e.g. one unambiguously documented by an expert, specimens identified and stored in a museum collection, etc.), while values close to 0 would be assigned to observations likely to be false. Consequently, total number of sightings (n) in Eq. 1 can be replaced by an index value, which represents a sum of probability values assigned to each of the observations in a sighting record:

$$r = \sum_{i=1}^{n} p_i. \tag{4}$$

The value of r represents the most likely number of observations, while the difference between the value of index r and the total number of sightings (n) represents a total observation record reliability. Note that when all observations within a sighting record are certain (i.e. when assigned likelihood values to each of the observations have a value of 1), r will be equal to n. Mean observation reliability is represented by r/n.

If the last observation (t_n) is likely to be false, the previous one (t_{n-1}) should be then considered as the end of the sighting record. The probability that an observation is the last one in a sighting record is equal to the probability that it is a true observation (p_i), multiplied with the probability that each of the consequent observations are false ($1 - p_i$):

$$p_i \times \prod_{a=i+1}^{n} (1 - p_a) \tag{5}$$

To account for this problem, the last observation (t_n) in Eq. 1 should be replaced with t_r, which represents the most likely endpoint of a sighting record:

$$t_r = \sum_{i=1}^{n} \left(t_i \times p_i \times \prod_{a=i+1}^{n} (1 - p_a) \right) \tag{6}$$

Note that when the reliability of the last observation is equal to 1, t_r is equal to t_n.The modified Solow (1993a) equation will then be:

$$p = \left(\frac{t_r}{T}\right)^r, \tag{7}$$

$$p = \left[\sum_{i=1}^{n} \left(t_i \times p_i \times \prod_{a=i+1}^{n} (1 - p_a) \right) \Big/ T \right]^{\sum_{i=1}^{n} p_i}. \tag{8}$$

The same modification can be applied to Eqs. 2 and 3, which estimate the probable extinction time and the upper bound (T_{CI}) of a $1 - \alpha$ confidence interval:

$$T_E = \frac{r+1}{r} t_r \tag{9}$$

$$T_{CI} = \frac{t_r}{\alpha^{(1/r)}}. \tag{10}$$

The main issue when applying the presented approach is to estimate and assign proper sighting reliability value to each of the observations. There is often a lack of precise data or criteria on the reliability of the observation record. However, this may be resolved by

applying different sets of probabilities. For example, the most and the least conservative sets of values could be applied, resulting in a range between the two predicted endpoints. Moreover, as discussed below, another solution would be to establish a standardized set of criteria regarding the reliability of different types of observations.

Sensitivity analysis

A sensitivity analysis was conducted to test the performance of the presented method under different sighting record scenarios, and to compare its behaviour with the standard use of Solow method (i.e., Eqs. 1–3). Sensitivity analysis was based on the general approach described by Rivadeneira et al. (2009). A series of artificial sighting records, populated by both false and true sightings, was generated and simulated in the R programming language (R v.2.12.1). Lengths of sighting time series were allowed to vary between 40 and 160 years. Sighting probability, as a proxy for the probability of species occurrence and detection probability, was allowed to vary between 0.2 and 0.4. Values for both of the ranges were drawn for each dataset from a uniform distribution. Two different extinction scenarios were applied, sudden extinction (with the sighting probability remaining constant over time) and a gradual extinction (with the declining sighting probability towards the end of the observation record).

Finally, for each of the sightings in a record, a value representing its reliability (i.e. probability that the sighting is true, p_i) was drawn from a uniform distribution, and consequently applied as a probability value in a binomial distribution to decide whether a certain sighting in a record is true or false. In this way, sightings characterized by a low reliability were more likely to be assigned as false, and vice versa (i.e. sightings with reliability value of 0.2 were on average assigned as true 20 % of the time, and 80 % of the time as false). The extinction event was randomly selected for each sighting record to occur anywhere between the middle and the end of the whole sighting record, and only false sightings were allowed to occur during the remaining part of the sighting record. Different scenarios were generated with three different ranges of the overall sighting reliability–low, medium and high (p_i values were allowed to range between 0 and 0.6, between 0.2 and 0.8, and between 0.4 and 1.0, respectively).

Sensitivity analysis was simulated with 10,000 iterations for each scenario. Each iteration was used to estimate the upper bound of a $1 - \alpha$ confidence interval (T_{CI}) using the presented approach (Eq. 10), as well as the standard Solow equation (Eq. 3). The Solow equation was applied with three different criteria regarding the acceptable level of sighting reliability: (1) "liberal criteria", where all observations were included in the analysis, disregarding their reliability, (2) "intermediate criteria", where only the observations with reliability falling within the upper 50 % of a given reliability range were included in the analysis, and (3) "conservative criteria", where only the observations falling within the uppermost 25 % of a given reliability range were used.

Estimated upper bounds of a 95 % confidence interval (T_{CI}) were evaluated based on their statistical coverage (i.e., on the appropriateness of the 95 % confidence intervals). An accurate and precise method would have 95 % of simulated extinctions falling within the upper bound (95 %) of the confidence interval (Rivadeneira et al. 2009). A method can be considered as more prone to a type I error (i.e. more likely to underestimate the true time of extinction of a species) if it produces the resulting coverage below 95 %. Conversely, those with the resulting coverage above 95 % can be considered as too conservative and therefore more prone to a type II error (i.e. they would be more likely to overestimate the true time of extinction). Difference between the true time of extinction and the estimated

upper bound of 95 % confidence intervals was also assessed (ΔT_{CI}). The extent to which ΔT_{CI} is influenced by different characteristics of a sighting record, such as the sighting record length, true time of extinction and the number of true and false sightings, was assessed by the use of Spearman nonparametric correlation test.

Additional set of sensitivity analysis scenarios was developed where the presented approach was applied on different methods. The evaluated methods, beside Solow approach (SOL), were S&S (Strauss and Sadler 1989), S&R (Solow and Roberts 2003) and MCY (McInerny et al. 2006). Each of the equations were modified so they would estimate upper bounds of a 95 % confidence interval (T_{CI}), as presented by Rivadeneira et al. (2009; for corrections to some of the equations see Jarić 2014). Consequently, the presented approach was applied on each of the equations by simply replacing t_n with t_r and n with r, as was presented above for Solow approach. In this set of scenarios, reliability of sightings was allowed to range between 0 and 1.

Example

As an illustration of the method application, the sighting records of the four potentially extinct bird species have been used (Roberts et al. 2010): the Eskimo Curlew, Ivory-billed Woodpecker, Nukupu‘u and O‘ahu ‘Alauahio. These sighting records are consisted of both reliable and controversial observations, and they were classified into three types of observations (Roberts et al. 2010): (1) physical evidence only, such as uncontroversial museum specimens, (2) independent expert opinion, and (3) controversial sightings. In order to assign specific reliabilities of each individual observation, we have applied a sighting reliability scoring system used by BirdLife International (Table 1; Lee et al. 2014). Based on this system, we classified sightings of the four assessed species as: (1) records described as being based on collected individuals (reliability interval 0.8–0.9), (2) records based on observations described in the literature as confirmed or considered fairly convincing (0.6–0.8), and (3) records described in the literature as, or judged to be, unconfirmed or questionable (0.1–0.4; Table 2). Consequently, the time of extinction of each species was estimated with three different sets of overall reliability levels applied—one with the upper endpoints of given reliability intervals (i.e., 0.9, 0.8 and 0.4 were applied for the three sighting categories), one with the mean values of the three reliability intervals (0.85, 0.7 and 0.25), and the one with the lower interval endpoints (0.8, 0.6 and 0.1). These three scenarios of data reliability estimations allow the production of a probable span of resulting predictions.

Results

Sensitivity analysis

Results of the sensitivity analysis are presented in Figs. 1 and 2. In scenarios with an instantaneous extinction, the modified method resulted in more conservative predictions than the three sighting reliability criteria applied to the standard Solow approach (Fig. 1a). The use of the standard Solow approach with conservative criteria was the least prone to make a type II error. Conversely, conservative criteria produced on average the longest confidence intervals, with the largest variation among iterations (Fig. 1b). In other words, conservative criteria required on average a longer period of time to pass before the extinction of a species could be inferred with certainty. An additional problem when using

Table 1 Sighting reliability scoring system by BirdLife International (Lee et al. 2014)

Sighting reliability	Observation type
0.90–0.94	Photograph or sound
0.8–0.9	Record described as being based on collected individual
0.6–0.8	Record based on observation described in the literature as 'confirmed' or considered fairly convincing
0.4–0.8	Record based on observation (with no qualification in the literature or with no additional information upon which to judge veracity)
0.1–0.4	Record described in the literature as (or judged to be) unconfirmed or questionable
0.1–0.3	Local report

Table 2 Sighting records for four possibly extinct species based on three different levels of sighting quality from Roberts et al. (2010) with corresponding sighting reliability based on the scoring system of BirdLife International (Lee et al. 2014)

Species (geographical population)	Sighting quality and reliability	Sighting record (years)
Eskimo Curlew *Numenius borealis*	Physical evidence (0.8–0.9)	1900, 1901, 1902, 1903, 1904, 1905, 1906, 1908, 1911, 1912, 1913, 1914, 1918, 1932, 1962, 1963
	Independent expert opinion (0.6–0.8)	1926, 1929, 1930, 1933, 1937, 1939, 1945, 1947, 1948, 1950, 1959, 1960, 1961
	Controversial sightings (0.1–0.4)	1956, 1964, 1968, 1970, 1972, 1973, 1974, 1976, 1977, 1980, 1981, 1982, 1983, 1985, 1987, 1990, 1992, 1996, 2002, 2006
Ivory-billed Woodpecker (U.S.) *Campephilus p. principalis*	Physical evidence (0.8–0.9)	1897, 1898, 1899, 1900, 1901, 1902, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1913, 1914, 1917, 1924, 1925, 1932, 1935, 1938, 1939
	Independent expert opinion (0.6–0.8)	1911, 1916, 1920, 1921, 1923, 1926, 1929, 1930, 1931, 1933, 1934, 1936, 1937, 1941, 1942, 1943, 1944
	Controversial sightings (0.1–0.4)	1946, 1948, 1949, 1950, 1951, 1952, 1955, 1958, 1959, 1962, 1966, 1967, 1968, 1969, 1971, 1972, 1973, 1974, 1976, 1981, 1982, 1985, 1986, 1987, 1988, 1999, 2004, 2005, 2006
Nukupu'u *Hemignathus lucidus*	Physical evidence (0.8–0.9)	1837, 1838, 1888, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1899
	Independent expert opinion (0.6–0.8)	1879
	Controversial sightings (0.1–0.4)	1960, 1961, 1965, 1967, 1968, 1973, 1974, 1975, 1976, 1978, 1979, 1981, 1983, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996
O'ahu 'Alauahio *Paroreomyza maculata*	Physical evidence (0.8–0.9)	1837, 1888, 1891, 1892, 1893, 1896, 1897, 1901, 1902, 1903, 1950, 1968
	Independent expert opinion (0.6–0.8)	1936, 1937, 1939, 1940, 1946, 1947, 1948, 1949
	Controversial sightings (0.1–0.4)	1952, 1955, 1956, 1957, 1958, 1960, 1961, 1964, 1966, 1968, 1969, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1984, 1985, 1989, 1990, 1991, 1993, 1996, 1997, 2000, 2001, 2002

conservative criteria is that they often tended to exclude too many sightings from a record to be able to make estimations of the extinction time (i.e. it was not possible to use conservative criteria in 12–22 % of iterations, depending on the type of scenario). At the same time, the new modified method and the standard Solow approach with the two lower criteria were still able to make reasonable predictions for such datasets. In scenarios with a gradual extinction, the new modified method provided a better coverage than the standard Solow approach, with a nearly perfect coverage in scenarios with a low overall sighting reliability (Fig. 1a).

In scenarios characterized by a low overall sighting reliability, the presented method produced predictions that were similar to those when the standard Solow approach with conservative criteria was applied (Fig. 1b). With the increase of the overall observation reliability, predictions of the presented method became more similar to the results produced by the intermediate criteria. At the same time, the presented modified method produced more uniform predictions (i.e. they had a lower variation among iterations) than the standard Solow approach with intermediate or conservative criteria.

Assessment of the influence of different characteristics of a sighting record on the resulting confidence intervals indicated that both the new method and the standard Solow approach were mostly sensitive to the number of false sightings recorded after the true time of extinction, as well as to the length of the sighting record after the true time of extinction (Spearman's Rank nonparametric correlation test; $p < 0.01$; Table 3). The sensitivity of the new method regarding these two parameters was similar to that of the standard Solow equation with liberal criteria applied. However, the new method was more sensitive to the true time of extinction and to the number of true sightings than the standard Solow approach. It was also more sensitive to the overall sighting probability and to the number of false sightings recorded before the true time of extinction.

Performance of the presented approach, applied on different methods (Fig. 2), indicated that it tended to provide more conservative predictions in scenarios with an instantaneous extinction than the standard Solow (1993a) and McInerny et al. (2006) methods with conservative or intermediate criteria. On the other hand, it produced similar confidence intervals as the two methods. In scenarios with a gradual extinction, the standard Solow (1993a) and McInerny et al. (2006) methods tended to be more prone to underestimating the true time of extinction, and the inclusion of the presented approach improved the coverage of the two methods (Fig. 2a). The method of McInerny et al. (2006) was especially prone to produce poor coverage for datasets with a gradual extinction, which was also observed by Rivadeneira et al. (2009).

The standard Solow and Roberts (2003) method was overly conservative when applied for datasets with an instantaneous extinction, and the inclusion of the new modification resulted in an improved coverage. On the other hand, the standard use of the Solow and Roberts (2003) method with conservative criteria provided better coverage for datasets with a gradual extinction than the new method, which tended to underestimate the true time of extinction. For both types of datasets, the new method considerably reduced the length of confidence intervals from those produced by the standard Solow and Roberts (2003) method (Fig. 2b).

The Strauss and Sadler (1989) method, applied with either one of the three sighting reliability criteria or with the new method, produced overly conservative predictions for scenarios with an instantaneous extinction. For scenarios with a gradual extinction, the new method provided better coverage than the standard Strauss and Sadler (1989) method with conservative or liberal criteria, while it was overly conservative when compared with the predictions produced by Strauss and Sadler (1989) method with intermediate criteria.

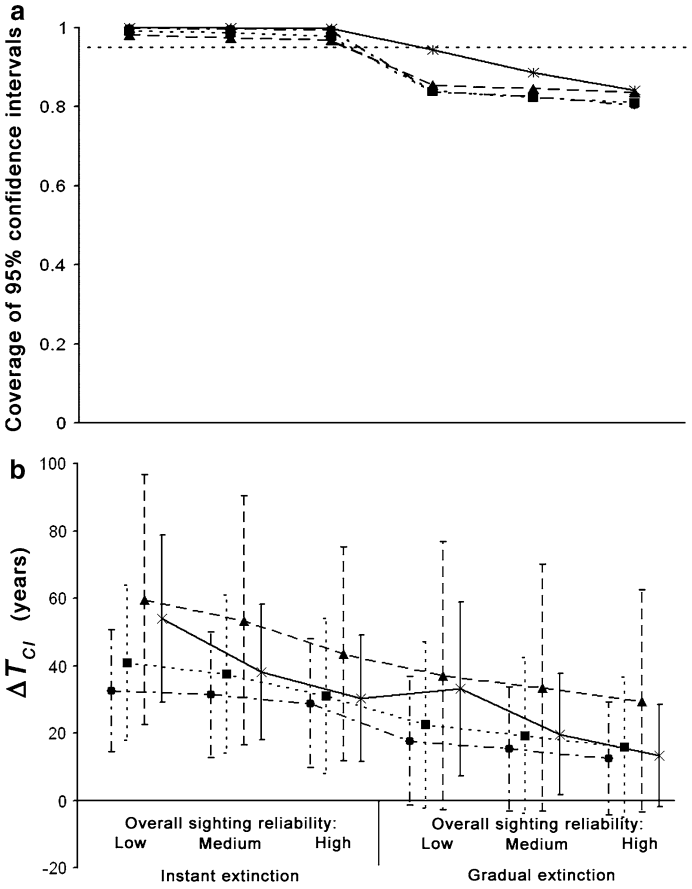

Fig. 1 Average estimations of the upper bound of a $1 - \alpha$ confidence interval (T_{CI}) for simulated sighting records with three different levels of the overall reliability of sightings (low, medium and high) and with two different modes of extinction (instant vs. gradual). *Solid line, stars* predictions produced by the new method (Eq. 10); the *three broken lines* represent predictions of the standard Solow equation (Eq. 3) with different criteria regarding the acceptable level of sighting reliability: *dash-dotted line, circles* liberal criteria (all observations included), *dotted line, squares* intermediate criteria, *dashed line, triangles* conservative criteria (see the text for explanation). **a** Coverage for 95 % confidence intervals under different simulation scenarios; the *broken line* shows the threshold value (95 %) indicating a perfect coverage (see Rivadeneira et al. 2009); **b** average difference (mean ± standard deviation) between the predicted 95 % confidence intervals and the true time of extinction (ΔT_{CI})

Again, the use of conservative criteria for any of the four methods could not be conducted for a significant portion of datasets (16–22 %), as it tended to result in too few sightings in a record to make predictions.

Example

Results indicated that Eskimo Curlew and Ivory-billed Woodpecker can be considered as extinct (Fig. 3). These results are consistent with those estimated by the standard Solow equation (cf. Roberts et al. 2010). With the overall increase of the sighting reliability

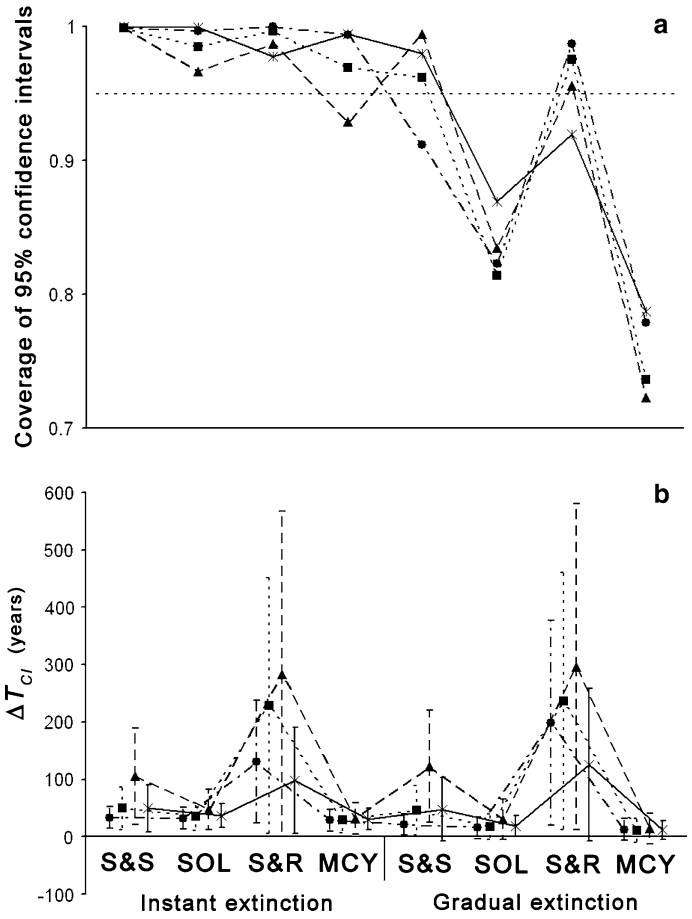

Fig. 2 Average estimations of the upper bound of a $1 - \alpha$ confidence interval (T_{CI}) for simulated sighting records of varying reliability and with two different modes of extinction (instant vs. gradual), based on the use of four methods: S&S (Strauss and Sadler 1989), SOL (Solow 1993a), S&R (Solow and Roberts 2003) and MCY (McInerny et al. 2006). *Solid line, stars* predictions produced by the new approach (Eq. 10); the *three broken lines* represent predictions of the respective method with different criteria regarding the acceptable level of sighting reliability: *dash-dotted line, circles* liberal criteria (all observations included), *dotted line, squares* intermediate criteria, *dashed line, triangles* conservative criteria (see the text for explanation). **a** Coverage for 95 % confidence intervals according to the four methods; the broken line shows the threshold value (95 %) indicating a perfect coverage (see Rivadeneira et al. 2009); **b** average difference (mean ± standard deviation) between the predicted 95 % confidence intervals and the true time of extinction (ΔT_{CI})

assigned, time of extinction and the upper bound of a 95 % confidence interval resulted in a delayed inferred extinction date, which is a different pattern than the one observed in the sensitivity analysis (Fig. 1b). Such differences are caused by the distribution of certain and controversial observations within the sighting period. In the sensitivity analysis scenarios, both types of observations were distributed randomly, while in sighting records of the four studied bird species, controversial sightings were situated mostly at the end of the sighting period. However, even with the application of the most liberal criteria (i.e. the highest

Table 3 Mean coefficients of correlation (Spearman nonparametric correlation test; $p < 0.01$) between different characteristics of a sighting record and the difference between the predicted 95 % confidence intervals and the true time of extinction (ΔT_{CI}), produced by the standard Solow approach (Solow 1993a) with three different criteria levels of sighting reliability and by the new method presented here

	Liberal criteria	Intermediate criteria	Conservative criteria	New method
True time of extinction	−0.15	−0.11	−0.04	−0.16
Time elapsed between the true time of extinction and the end of the whole sighting record	0.77	0.48	0.28	0.65
Overall sighting probability	−0.08	−0.16	−0.19	−0.20
Number of true sightings	−0.07	−0.17	−0.14	−0.21
Number of false sightings recorded before the true time of extinction	−0.07	−0.11	−0.07	−0.18
Number of false sightings recorded after the true time of extinction	0.83	0.51	0.27	0.71

overall data reliability assigned), estimated time of extinction of both species would occur no later than in 2010 (T_{CI}).

Predictions were more conservative regarding the other two species. Results suggest that the Nukupu‘u and O‘ahu ‘Alauahio might still be extant, as the upper bounds of 95 % confidence intervals were found to be 2020–2027 and 2018–2021, respectively. On the other hand, the likely time of extinction for the two species was estimated to correspond to the years 1995–2003 and 1997–2006, respectively, supporting the hypothesis that they are extinct. Results for these two species were more similar to those produced by the standard Solow equation when controversial sightings are included in the analysis (Roberts et al. 2010).

Discussion

The novel approach presented here was developed as a modification of the existing methods to account for the sighting reliability. In a similar way, modifications of the standard Solow method by McCarthy (1998) and McInerny et al. (2006) accounted respectively for sighting effort and sighting rate within an observation record. Although the presented approach incorporates sighting reliability in a way that can to a certain extent be considered *ad hoc*, its predictive power in estimating extinction likelihood was demonstrated through simulations, using the approach of Rivadeneira et al. (2009). Sensitivity analysis indicated that the method tends to be conservative, resulting in later predicted extinction dates than those of the standard methods. As such, it would appear to be less prone to make a type I error (i.e. to underestimate the true time of extinction). This is contrary to the majority of the existing methods, which are more likely to make a type I error (Rivadeneira et al. 2009). According to Butchart et al. (2006), precautionary approach is sensible, in order to encourage continuing conservation efforts until there is no reasonable doubt that the species is truly extinct, as advocated by the IUCN Red List criteria (IUCN 2001).

The presented method suggests that the Ivory-billed Woodpecker may be considered extinct. Such results are consistent with the findings of other authors (McKelvey et al. 2008; Elphick et al. 2010; Roberts et al. 2010; Gotelli et al. 2012; Solow et al. 2012). On

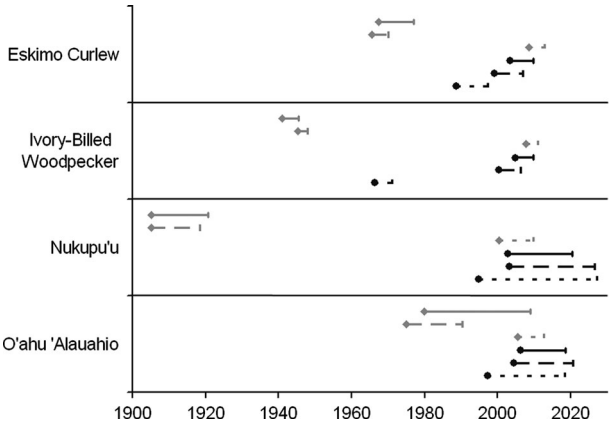

Fig. 3 Estimated year of extinction (*filled diamond* or *filled circle*) and the upper bound of a $1 - \alpha$ confidence interval (right-hand side of each bar) for the four potentially extinct bird species (sighting data from Roberts et al. 2010). *Grey bars* represent predictions of the standard Solow equation (see Roberts et al. 2010): *solid line* (*top* for each species) estimates based on the physical evidence data only, *dashed line* (*middle*) estimates when the independent expert opinion data are also included, *dotted line* (*bottom*) estimates when controversial sightings are included. *Black bars* represent predictions of the presented approach with different overall observation reliability levels assigned (see the text for explanation): *solid line* (*top*) high observation reliability, *dashed line* (*middle*) intermediate observation reliability, *dotted line* (*bottom*) low observation reliability

the other hand, Elphick et al. (2010) omitted all controversial observations from the analysis (i.e. those recorded after 1944), while Rout et al. (2010) estimated that the inclusion of controversial observations would make the Ivory-billed Woodpecker highly likely to be extant, with the probable time of extinction occurring even as late as in 2032. Such findings differ from the results of the presented approach, where the species was categorized as extinct even with recent controversial sightings included in the analysis. Moreover, as shown in Fig. 3, if the recent observations of the Ivory-Billed Woodpecker are considered to be of poor reliability, the date of extinction will shift to a much earlier date (i.e. the 1960s). This illustrates the sensitivity of the method to the reliability of observations, as well as to their distribution within the sighting record. Instead of applying some arbitrary threshold for the sighting reliability that is considered acceptable, the approach presented here incorporates the reliability of each observation into the resulting predictions, by estimating the most likely number of observations and the most likely time of the last observation. This is the major advantage of the presented method over those that do not account for observation reliability.

According to the results, the Eskimo Curlew may be considered extinct, while the Nukupu‘u and O‘ahu ‘Alauahio may still be extant, although with a low probability of persistence. Previous studies estimated that all three species are likely to be extinct (Butchart et al. 2006; Elphick et al. 2010; Roberts et al. 2010). However, this can be partially explained by different data used in those studies. Elphick et al. (2010) used only the non-controversial observations, while Butchart et al. (2006) applied a qualitative assessment framework that employs various types of data to infer extinction.

The standard Solow method assumes that there is no systematic trend in the likelihood of detection over time (Roberts et al. 2010). In case of sighting records that change in frequency over time, methods that account for this feature could be applied instead (Solow

1993b; Jarić and Ebenhard 2010). However, due to the high overall sighting effort and sighting probability in the four analyzed sighting records (Roberts et al. 2010), potential violations of the assumed underlying distribution of sightings are not likely to increase the risk of a type I error.

Similarly as it was demonstrated with the example of the standard Solow approach, other previously developed methods can also be easily modified, so that they incorporate considerations regarding the data reliability. As indicated in the sensitivity analysis, the same approach can be also applied to the methods developed by Strauss and Sadler (1989) and McInerny et al. (2006), by simply replacing t_n and n in the equation with t_r and r, respectively (Eqs. 4 and 6).

While the new method presented here does not replace the previously developed methods, it has the greatest potential when used as an additional method for sighting records that contain controversial observations. Nevertheless, a priori analysis of characteristics of the species or population in question and its sighting record might indicate which method offers the best potential to produce reliable results (Jarić and Ebenhard 2010).

The major issue of the approach presented here is the difficulty in estimating the reliability of each of the observations within a record, since it is a complex and largely subjective task. As stated by Roberts et al. (2010), each sighting that goes into determining whether the species is possibly extinct is still judged in a largely *ad hoc* fashion. Nevertheless, the proposed method may still be preferable to the standard approach of simply accepting or rejecting anecdotal data based on some arbitrary threshold value. In the present study, we applied the sighting reliability scoring system that is used by BirdLife International (Lee et al. 2014). While being generally simple, it still appears to be sufficiently robust to be applicable for preliminary sighting reliability evaluation. Nevertheless, future research should focus on developing more rigorous scoring systems, which would involve other characteristics of each record, such as experience and credibility of the observer (in case of unconfirmed reports), species prior abundance and distribution, observation history, presence of similar species, and expert judgement. The use of structured elicitation approaches for reaching consensus among experts, such as the one developed by McBride et al. (2012) should be also considered. Beside the presented scoring system, other classifications of observations could be also applied, such as the one developed by McKelvey et al. (2008). Within the study, the authors proposed the use of the following classification of evidentiary standards for occurrence records (ordered from the least to the most reliable type of records; McKelvey et al. 2008): (1) visual observations of organism's sign, (2) visual or auditory observations, (3) photograph, video, or recorded auditory evidence lacking diagnostic characteristics, (4) diagnostic track impressions, (5) diagnostic recorded auditory evidence, (6) photograph or video showing diagnostic characteristics, (7) diagnostic DNA evidence, (8) live or dead specimen. They also advocated the development of rules to standardize reliability ratings, which would also consider a species' rarity, prior evidence of its existence, and the goals of the study or survey (McKelvey et al. 2008). Such set of standards could also assign explicit probability values to indicate reliability of each type of observations under different characteristics of species and sighting records, which could be consequently used as input values for the proposed approach. Moreover, previous studies also estimated the likelihood of a recent controversial observation based on preceding series of sightings (Solow et al. 2006; Roberts et al. 2010). Such an approach could also be used as an additional tool for the presented method, to evaluate validity of controversial sightings and to infer their specific reliability.

The development of a more robust sighting reliability system should not be confused with the development of an overall system for assessing the probability of extinction or persistence of a species. For the latter, several factors come into play including (a) time since last sighting, (b) data availability/search effort, (c) type of population decline, (d) species detectability and (e) habitat availability. A number of these are covered to a certain extent in the method developed in this paper. However, habitat availability, search effort and knowledge of population decline require further study.

There is a need to develop a formal method of weighing different types of sighting evidence, which could be consequently incorporated into the IUCN guidelines to extend the consistency of species conservation assessments (Roberts et al. 2010). The presented method could represent a good step in that direction, by providing a meaningful way to incorporate the quality of sighting data in extinction likelihood predictions. Limitations of the existing data must be recognized and accounted for in policy decisions, even at the cost of basing decisions on imperfect data sets (Elphick et al. 2010). That represents the major advantage of the method presented here, as it recognizes and incorporates the reliability of each observation within a record and uses them to infer the likelihood and the time of extinction. This also allows for better comparability among different datasets than the standard approach of making arbitrary decisions whether to include or exclude controversial observations from the analysis, especially when confronted with datasets that are largely consisted of controversial sightings. Here we present a method that is generally simple when compared with other more advanced approaches (e.g. Solow et al. 2012; Lee et al. 2014). As stated by Rout et al. (2010), simplicity of a method also increases its transparency and its accessibility as a decision tool. This makes it easy to use and apply for the assessment of a larger number of species or populations.

Acknowledgments We acknowledge the support by the Project No. 173045, funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia. We would like to thank anonymous reviewers for providing helpful comments and suggestions that have improved the quality of the paper.

References

Burgman MA, Grimson RC, Ferson S (1995) Inferring threat from scientific collections. Conserv Biol 9:923–928

Butchart SHM, Stattersfield AJ, Brooks TM (2006) Going or gone: defining 'possibly extinct' species to give a truer picture of recent extinctions. Bull Br Ornithol Club 126:7–24

Collar NJ (1998) Extinction by assumption; or, the Romeo error on Cebu. Oryx 32(4):239–244

Elphick CS, Roberts DL, Reed JM (2010) Estimated dates of recent extinctions for North American and Hawaiian birds. Biol Conserv 143(3):617–624

Fisher DO, Blomberg SP (2011) Correlates of rediscovery and the detectability of extinction in mammals. Proc R Soc Lond B Biol Sci 278(1708):1090–1097

Fitzpatrick JW, Lammertink M, Luneau MD Jr et al (2005) Ivory-billed Woodpecker (*Campephilus principalis*) persists in continental North America. Science 308(5727):1460–1462

Gotelli NJ, Chao A, Colwell RK, Hwang WH, Graves GR (2012) Specimen-based modeling, stopping rules, and the extinction of the Ivory-Billed Woodpecker. Conserv Biol 26(1):47–56

Grogan CS, Boreman J (1998) Estimating the probability that historical populations of fish species are extirpated. N Am J Fish Manag 18:522–529

Hill GE, Mennill DJ, Rolek BW, Hicks TL, Swiston KA (2006) Evidence suggesting that Ivory-billed Woodpeckers (*Campephilus principalis*) exist in Florida. Avian Conserv Ecol 1(3):2

IUCN (2001) IUCN Red List Categories and Criteria, Version 3.1. IUCN Species Survival Commission, Gland, Switzerland and Cambridge, UK

Jarić I (2014) The use of sighting records to infer species extinctions: comment. Ecology 95(1):238

Jarić I, Ebenhard T (2010) A method for inferring extinction based on sighting records that change in frequency over time. Wildl Biol 16(3):267–275
Lee TE, McCarthy MA, Wintle BA, Bode M, Roberts DL, Burgman MA (2014) Inferring extinctions from sighting records of variable reliability. J Appl Ecol 51:251–258
Marshall CR (1990) Confidence intervals on stratigraphic ranges. Paleobiology 16:1–10
McBride MF, Garnett ST, Szabo JK, Burbidge AH, Butchart SHM, Christidis L, Dutson G, Ford HA, Loyn RH, Watson DM, Burgman MA (2012) Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email. Method Ecol Evol 3(5):906–920
McCarthy MA (1998) Identifying declining and threatened species with museum data. Biol Conserv 83:9–17
McInerny GJ, Roberts DL, Davy AJ, Cribb PJ (2006) Significance of sighting rate in inferring extinction and threat. Conserv Biol 20:562–567
McKelvey KS, Aubry KB, Schwartz MK (2008) Using anecdotal occurrence data for rare or elusive species: the illusion of reality and a call for evidentiary standards. Bioscience 58(6):549–555
Regan HM, Colyvan M, Burgman MA (2000) A proposal for fuzzy International Union for the Conservation of Nature (IUCN) categories and criteria. Biol Conserv 92:101–108
Rivadeneira MM, Hunt G, Roy K (2009) The use of sighting records to infer species extinctions: an evaluation of different methods. Ecology 90:1291–1300
Robbirt KM, Roberts DL, Hawkins JA (2006) Comparing IUCN and probabilistic assessments of threat: do IUCN red list criteria conflate rarity and threat? Biodivers Conserv 15:1903–1912
Roberts DL, Kitchener AC (2006) Inferring extinction from biological records: were we too quick to write off Miss Waldron's Red Colobus Monkey (*Piliocolobus badius waldronae*)? Biol Conserv 128:285–287
Roberts DL, Solow AR (2003) When did the dodo become extinct? Nature 426:245
Roberts DL, Elphick CS, Reed JM (2010) Identifying anomalous reports of putatively extinct species and why it matters. Conserv Biol 24(1):189–196
Rout TM, Heinze D, McCarthy MA (2010) Optimal allocation of conservation resources to species that may be extinct. Conserv Biol 24(4):1111–1118
Sibley DA, Bevier LR, Patten MA, Elphick CS (2006) Comment on ''Ivory-billed woodpecker (*Campephilus principalis*) persists in continental North America". Science 311:1555a
Solow AR (1993a) Inferring extinction from sighting data. Ecology 74:962–964
Solow AR (1993b) Inferring extinction in a declining population. J Math Biol 32:79–82
Solow AR (2005) Inferring threat from a sighting record. Math Biosci 195:47–55
Solow AR, Roberts DL (2003) A nonparametric test for extinction based on a sighting record. Ecology 84:1329–1332
Solow AR, Kitchener AC, Roberts DL, Birks JD (2006) Rediscovery of the Scottish polecat, *Mustela putorius*: survival or reintroduction? Biol Conserv 128(4):574–575
Solow A, Smith W, Burgman M, Rout T, Wintle B, Roberts D (2012) Uncertain sightings and the extinction of the Ivory-Billed Woodpecker. Conserv Biol 26(1):180–184
Strauss D, Sadler PM (1989) Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Math Geol 21:411–427
Thompson CJ, Lee TE, Stone L, McCarthy MA, Burgman MA (2013) Inferring extinction risks from sighting records. J Theor Biol 338:16–22